

# **National Registry of Effective Programs (NREP) Substance Abuse Treatment**

**Hilton Embassy Row Hotel  
Washington, DC  
December 9–10, 2003**

## ***Welcome and Opening Remarks—Andrea Kopstein***

This is the kick-off meeting for expanding the NREP process to Substance Abuse (SA) Treatment (currently only in the prevention arena). It is important to catalog treatment approaches that produce good results consistently. NREP gives good way of linking practice to service. Will be modifying the current system to make it fit for SA treatment. NREP Criteria will identify treatment programs to move forward.

Mady Chalk: This project is really important. Need to have some criteria to tell what works best in treatment, so today is a real opportunity to produce these criteria to judge what an effective program is. There are a host of issues—how tightly or loosely we draw this, what we mean by treatment program. This is something we need conversation about; we need to start a dialogue so we can implement a program. Interplays with a lot of other things going on. What does performance have to do with an effective program?

## ***Overview of the Agenda—Andrea Kopstein***

Introduced Tom McLellan (facilitator for this meeting). Participants introduced themselves.

Paul Brounstein: Idea behind NREP was to provide a resource for communities, not to be prescriptive. To create an engine that drives a translation—is there good science driving this program or strategy. If so, what are the results? Sometimes something has a null or negative effect. It is important to help communities develop something to implement these activities. Coding (rather than rating) criteria describe the treatment—how user friendly, resources available, etc. Registry itself lists these programs that have been shown to work with specific populations under specific circumstances = critical components. A key requirement is providing technical assistance and resources. How does the community adapt for its own population. Adaptation is as yet an inexact science. Work is ongoing for fidelity and adaptation. Cognitive behavioral therapy is a practice, not a program. What is gained by looking at program examples that did not work? Need to focus on whole set of interventions that form a program. Each community will have to adapt these programs for itself.

NREP is one response to SAMHSA providing service. It is not the answer; it is a piece of a system.

Mady Chalk: Funding replication of programs in the CSAP registry for the first time this year. Need to determine how we define a program and what it takes to make a program effective.

***Brief History of Evidence-Based Practice (EBP)—Steve Leff (HSRI)***

Science to services is anarchic, market-driven. This equals conflicting messages. We should bring evidence together; each trial is a piece of the whole.

We know intervention science is a phased process; need to test for safety and effectiveness; multiple studies for each phase; stakeholder participation—stakeholders should be the arbiters of where we are going. Evidence should be the means and not the end. Need a supporting organization, a structure to provide continuity for the development of evidence.

Scientific concerns are what we are here to discuss: e.g., appropriate controls (p5 M).

***Reasons and Goals for NREP Extension to SA Treatment—Mady Chalk***

What is NREP? The NREP process is a resource for reviewing and identifying effective evidence-based programs. NREP was begun by the Center for Substance Abuse Prevention (CSAP) (p1).

NREP classifications (model, effective, promising, insufficient current support programs) (p2).

Review process in two phases (p2); refinements include integrity and utility.

Summarized current status with programs and grant applicants.

Overview of NREP and its potential for expansion to treatment.

We want to accelerate application of interventions and make sure that they are appropriate and high quality; now there is a huge time lag.

Science to Service Initiative involves research and development, dissemination and implementation (community infrastructure needs to exist), monitoring and feedback. SAMHSA has been redesigning its discretionary grants—four initiatives.

Overarching intent of NREP expansion is to make sure that info is disseminated to people who need it (p.9—Brooks).

***Facilitator's Perspective on the Process of the Meeting—Thomas McLellan***

Brounstein: One reason for the surprising numbers is that States require some money be spent on evidence-based programs (science-based); States started seeing return on investment. Block grants started using, so States now use that wording. State incentive grants (5-year, \$3 million) show that effects produced are twice to four times that of the baseline.

## McLellan: Development of Practices

What shapes practices? Process research; best practices statements and guidelines; licensing standards; all tell providers what to do. Practicality and cost constraints also matter. Done by researchers and providers—tell people what to do.

What develops outcome expectations? What do you want it to do? We want a cure, but success is evaluated as the number of people who have sustained abstinence after treatment. Expectations are related to symptom management (mental health [MH] approach). Managed care has expectation about cost containment. Want to contain social harms. Sometimes expectations conflict with practices. E.g., in Delaware they do not understand the problem, but they want people to stay in treatment and they want no more arrests and they do not care how this is achieved. Segregated funding for SA; ambiguity about SA as a mental health illness or just a bad habit. No professional schools for addiction; addicted community largely determines what goes on, but they are invisible—influence, but no political clout. New things are typically mandated, not requested by consumers. This is a terrible time for the SA treatment field—only area where insurance claims are going down; no constituency is making demands. Tremendous workforce turn over; no information systems. Result: “programs” are very similar around the country but are small (~60 percent not for profit), ~85 percent outpatient, short-term.

Matrix Program—set of things to be delivered, e.g. methadone programs; drug courts; Minnesota model (residential program); therapeutic communities (was 12 months, now 30 to 90 days).

If it does not fit, it will not happen. If there is no cost, it will not fit and it is difficult to make SA fit.

The cloak of evidence-based treatment comes from having research funding from the National Institute on Drug Abuse (NIDA) et al., which gives researchers a chance to show their wares. E.g., motivational enhancement therapy, but is it better than keeping clinic open evening hours—will never get a grant to find that out.

SA and MH are combined, but have different histories; cannot just plug SA into MH solutions.

How is the system structured to deliver care?

What is the fit? What has to change?

## *Meeting Participants' Experience With NREP*

### **Frank McCorry**

Explained the grant review process. Mr. Brounstein's characterization that Mr. McLellan picked up on: What are we trying to create? Difference between being a resource and the list that gets disseminated, so people play to the list. The Centers for Disease Control and Prevention (CDC)'s HIV prevention program for adolescents is a good exemplary practice. They learned there were

problems with a lack of family involvement. Second intervention was different because it focused on family involvement. Whole issue of *replication* with fidelity is important.

*Sustainability* of effects—what is the standard? How do we determine what constitutes effectiveness in an area where there is a substantial dropoff post intervention?

*Effectiveness on drugs and alcohol use*—real-world application. No matter how well controlled the program, if cannot work in real-world (troubled, four diagnoses) population, it is not effective.

Leads to a pool of resources that practitioners can access.

### **Flo Stein**

High expectations around the country regarding EBP. Process is so critical that we are starting to implement and States have started to fund it. North Carolina (NC) is a practice-improvement collaborative State. Wanted to use program as an infrastructure development project and wanted to gain control over their moneys. First program showed harm to the children treated. Models are very well bounded; treatment is harder to bound. Outcomes got better. Learned that to prevent long-term problems, there are interventions that work, but only for certain children. Decided to identify population, interventions, and outcomes and will pay only for those. Have best practice documents. Trying to shape the behavior of their providers.

Model versus intervention, but inside are EBP.

Consumers, providers, and purchases need a better way to think about these things. NC is defining its system very tightly and will loosen as they learn. The data they get is not predictable and will be difficult to interpret. This is not real science. It is time to stop paying for things that are not helpful. However, NC has lost no funds.

### **Roy Gabriel**

A good thing about NREP is not insisting on clinical trials, but in bringing together all kinds of evidence.

Researchers should work hand in hand with practitioners = cultural change. Dialogue is meant to be bidirectional. Not defensible to insist upon adhering precisely to a model, but must accommodate to particular population, etc.

Concerns about effort: it may degenerate to “being on the list.” Focus on implementation, not fill in a checklist.

Practitioners in OR identify four or five treatment models.

### ***Overview of NREP Criteria and Role of Reviewers—Steve Schinke***

NREP is first step in overall dissemination. Bulk of activity is spent in getting information out in various ways (hard copy, Web site, block grants, etc.). Getting feedback to complete the loop. NREP topics—growing list.

“If you build it they will come”: Once people see the pay-offs, programs will come into the process, as they have for each topic.

Best programs to come into the system (often community, grass-roots organizations) that have capacity and desire to ramp up. Often university researchers cannot or do not want to.

In next generation NREP should look for new data.

At review level, they are only interested in outcome data.

Reviewers largely in social sciences, 44 percent women, 30 percent various minorities. Described reviewer selection and training—now use a Web-based, telephone system, instead of training meetings. Described workload, turnover. Paid \$43.75 per hour.

Important part of the process is providing technical assistance to people who need help in submitting applications to be reviewed.

21 NREP rating criteria, largely methodological; integrity and utility are the real decision-making criteria.

Interrater reliability is important.

Our task this afternoon and tomorrow is to parse out the 21 criteria and determine whether they are appropriate for SA.

### ***Discussion***

- Explained rating and scoring process; it is not an aggregate or summary, analogous to NIH process of priority score that represents everything that went into the review. It is an overall assessment, not a summary.
- Effective score (5.0–4.0) goes to Model Programs. Reason for model is that it is something you can replicate. Have been able to talk to developer to assist with implementation and training. Emphasis is on adaptation to a particular community.
- Must develop a second level of TOT in State. Cannot develop enough with frequency to make it affordable.
- Information a potential implementer received gives population it was tested on; they can apply it as they will. With drugs, there is an off-label use. What does it mean to be effective?
- Protocol for types of material people submit for review? Prefer manuscripts, preprints, reprints, science reports. (Receive more and more applications geared specifically to NREP.) Want programmatic materials, whatever they use to actually implement the program. Look at

data first, and then invite all the backup material. The burden is on the developer to provide all the material necessary. There also is an appeal process.

- A structure is in place to support some activities, but not others. CSAP funded National Center for Advancement of Prevention. In New York they identify programs (not disseminate). Dissemination went to Northrop Grumman. SA structure needs support mechanism but does not have to be supported in exactly the same way as other areas.
- System to monitor programs? Do they continue to have the dissemination capacity? Does field experience same results as in the lab? Next generation of NREP is to rereview those programs.
- There is no formal decision tree; these scales are meant to guide reviewers and provide feedback to submitters.
- Wide discrepancies among reviewers <20 percent.
- Integrity is a measure of the reviewers' confidence that the investigators established a causal relationship between the treatment and the outcome.
- Use same criteria, but what the client agency does with the material differs.
- Cost data are developed and disseminated on the Web. It is information they get from the developer; document outcomes, how implement program and what the requirements are to do that (staff members, other resources, cost). Will develop budget sheet the potential user can adapt. Effect size is also made available.
- Comcast and Consumer Reports have approached Tom McLellan about this exact project. NIDA has done this; we do not have to invent it.

### ***Review Criteria Regarding Extending NREP to SA Treatment—Steve Schinke and Tom McLellan***

NREP Criteria Crosswalk, i.e. NREP evaluative criteria were compared with descriptive criteria, some of which have changed. This is not a consensus process; just brainstorming.

Are criteria weighted? Number of comments also is significant.

#### **1. Theory/Conceptual Underpinnings/Hypothesis**

- Why is theory important? Seems to be something that occurs after you get a result. Logic model may be a little closer to what people have in mind. It is something people found to work. Most medications have unexpected results, so the idea that you have to start from a theoretical framework does not fit.
- Communicates something different when you introduce theory: connotes approved Federal list.
- Lack of empirical relationship between this criterion, theory, and the outcome measures. See no reason to keep theory as part of the matrix.
- Logic models are the conceptual framework. Treatment is thought of as a single, isolated item; it is important to change this notion. It would be nice to base discussions here on what CSAP has done already. A lot of what treatment is about has to do with stages of treatment. Performance measures need to relate to empirically driven actions. What contributes to early recover? What leads to retention? All lead to better outcomes in the long run. Must look at interventions as being purposeful.
- Association of target behavior and the intervention that is intended to change the behavior.

- Programs are really funding entities. Interventions are building blocks that happen in those programs. To complicate it, there are multiple levels of interventions (#1 is simple, e.g., anger management; #2 is modules, a series of interventions; #3 is a strategy). Within that context is organizational behavior—how do you characterize organizations and their intervention programs; has to do with whether they are ready to adopt new strategies. This is important to define what we mean by treatment. What are they intended for? Who are they intended for? There are specific stages to treatment for which interventions are designed. Most of what happens in treatment is outpatient, short-term.
- People need resources in a virtual library for whatever aspect of treatment they are doing. Should be based on outcomes, not 5-year followup studies. No one does a single sterile intervention. Interim measures should be part of the design criteria, so they will link together over time.
- There is an a priori link between the procedure, intervention and at least one result or targeted behavior.
- Is it incumbent upon the reviewer to place the program in the framework's context?
- Can have a good program and be ignorant of the entire framework. May not be able to articulate the theory, but have a good intervention.
- It is critical that the evaluators have a clear idea of where the project fits in.
- National Treatment Plan and NIDA's document are reference documents. Does a reviewer import that?
- Data do not come from the theory, but from the content area. "Intervention addresses content areas that have an evidence-based association with targeted areas or outcomes." This wording gets away from being judgmental; articulation is part of the dissemination process.
- Theory may help counselors, who are not academically trained, understand how all these things hang together.

## **New 2. Study Population**

- Even more so in SA, we need an accurate description of the study population. For other research it would be important. Age, ethnicity, coerced into treatment?
- How much do you discourage exportability by being very specific about the population?

## **2. Intervention Fidelity**

- No comments.

## **3. Process Evaluation**

- What would it give you beyond fidelity? That is what previous groups said.
- The issue is to go beyond intervention fidelity. Process evaluation is critical for implementation—lessons learned.
- May think about the potential of general application with fidelity.
- Threaten validity by giving alternatives to the outcomes
- Does not drive journal article acceptance. Question of experience informing other applications is important.
- Full-fledged interventions for outpatient programs.

#### **4. Design**

- Should be written for provider as well as the researcher, so people who are not researchers understand it.
- FDA uses “active control or incremental control.” Could be useful for drug-abuse treatment too. Must be better than placebo is there is no effective intervention, but if there are already competitive interventions, then it is as good to show that your intervention is at least as good. In the real-world of SA, 99 percent of interventions have never been tested. Testing against placebo is also a thorny issue on ethics grounds.
- Type 1/type 2 error problem is interesting. Why would they be penalized for showing no effectiveness? When talking about comparison conditions, helps to know what criteria you’re measuring against.

#### **5. Analysis of Effects**

- If biases adequately addressed, then subject recruitment is less an issue. Verbiage talks about bias re the target population, but have they stated who the target population is?
- Random assignment is really tough to do. Using Medicaid dollars, it is challenging to get permission to do random assignment because people are reluctant to appear to be withholding care, even though they would give the usual standard of care. “Partially randomized” covers a lot of sins—should drop it.

#### **6. Analysis: Sample Size**

- Type 2 concerns apply. Is it necessary if programs are assumed to have an effect?
- Rather than power, what about number of tests conducted?

#### **7. Attrition**

Deleted—not interested in whether, but how it was handled.

#### **8. Analyses of Attrition Effects**

- This is arguably one of the most debated topics in statistics. Are you talking about missing data, failure to locate, or dropout? Subjects lost to the study—could be either of the last two. It is a technical issue.
- However, you cannot analyze data without dealing with people who failed to return. If people do not find the treatment attractive, it is one kind of outcome. If you cannot find them or if they failed to answer that question. They are all different situations and need to be accounted for differently. This assumes at least partial treatment. This criterion must be divided into at least two.
- This was discussed extensively.
- Some people recruit many people, but only a few are enrolled in the study. How do you account for this? Do not know how the sample is biased by the people who refuse to participate. As a standard, the FDA will not consider a study without at least 70 percent acceptance.
- There might be a criterion on palatability of care to capture this.

#### **10. Outcome Measures: Substantive Relevance**

- So many in SA expect to have an effect on the target measure, but in fact affect something like employment or housing.

- The outcome measures must have relevance to the stated issue. The issues is what did you intend to affect. Also need something that grounds it more specifically that targets it into a phase of care. This is a bit free-floating.
- Unanticipated effects that are not conceptually related might be rich and valuable; do not want to tell people we do not want to hear about anything they did not plan for.
- One way to address a prior identification is the FDA's: they require picking a primary measure and a few secondary measures.
- Would you not want to look at a number of measures and have a convergence?
- Prevention is risk-factor driven; you are looking for more that are concurrently achievable.
- The conceptual framework is usually loose enough to back-in any finding. Outcome identification may be important to researchers, but not policymakers.
- Mediating, having to do with the logical sequence of thing.
- If you see interim changes also born out in longer term, that is good.
- Capitalizing on chance, you could make data show what you wanted it to.

### **11. Outcome Measures: Psychometric Properties**

- Some outcomes we get in juvenile justice are not usually reported in studies. Some measures do not meet reliability and validity measures, e.g., self-reported drug use.
- With kids, people are looking at operational results—are they at school or at home when they are supposed to be?
- The public at large (e.g., Congress) does not buy self-reporting.
- Maybe a way is to say “consideration was given to inherent error.”
- A related issue is differences in how the variables are defined. Might find vertical integrity by a particular researcher, but definitions throughout the field might not be the same.

### **12. Missing Data**

- Combined with #13 in the revision.
- In SA, is there an issue of when the data are missing?
- Dropout, failure to contact and within contact. Missing data should be the three categories. Does not matter where the missing data occur.

### **13. Report on Missing Data**

Deleted

### **14. Outcome Data Collection**

- Demand characteristics of who is collecting the data.
- Would like to see a criterion about the quality of the outcome data reported—in person, independently collected, self-reported, etc.
- Quality of a process around data collection, which is different. The outcomes we are interested in are differences between groups. Evaluative criteria on left read fine.
- High attrition rates make it difficult to conduct research over time. Does time have weight? A continuing intervention should be considered differently than a more focused intervention, so it should be time appropriate. This could be an underlying issue: what is exemplary in a field with a relapsing condition? What we define as success gets to the heart of the issue of where you look for success. So if there is a lot of fall-off between end of study and followup, was it successful? Ability to reach beyond the contact episode is desirable, but is it a criterion for

success? E.g., Heroin treatment is effective immediately after treatment ends, but 6 months later they are all addicted again—was the treatment effective?

- Is the goal sustainability? What do they want? Short-term suspension or long-term suspension?
- Think about what this can do. If we can hint at this when we put this out, people will stop presenting themselves in general and vague ways that do not make sense (e.g., I am a heroin detox program, rather than I am a treatment program).
- Disconnect between how we promote EBP and the “what” the service is designed to accomplish. Promote EBP as things that will be effective forever, but then...
- Truth in labeling is the issue here.
- Take away depression treatment and the symptoms come back. Not so with SA.
- Could move the field ahead if we could move ahead incrementally rather than assume the improvement will last forever.
- It is essential that EBP does not become oversold in a way that prevents people from becoming educated about the condition. Cannot guarantee where a subject will be in 6 months.
- In comparison, in cancer care they do not consider number of relapses in whether the treatment was successful.

## **15. Analysis**

- Should add another threat to validity: respondent bias, lack of blinding. In SA, one is particularly relevant—time at risk at followup. Can get beautiful random assignment, beautiful followup, but it is because one group is not at risk of SA because of threat of incarceration or something else. If measure is drug use in 30 days and entire group wound up in jail, the rate is great.
- Violation of assumptions? Categorical data versus specific. Another aspect is richness of the analysis that takes place—are they looking at experiment versus control or length of stay, gender, age, multiple methods, etc.?
- Sufficiency of analysis might be better. But would not want to rule out all the tests; they might be the most appropriate.
- Plausible threats to validity—no comment.

## **16. Other Plausible Threats to Validity**

- Time at risk should be considered. Pertains to measurement instrument design. Validity analysis is much narrower.
- Blinding in psychosocial treatment does not work.

## **17. Integrity**

- Is this a summary measure? If so, you need a weighting mechanism for all the others.
- This is qualitative—overall level of confidence.
- What does this add?
- This is not a measure of integrity, but of confidence.
- And, it is equal to utility.
- If you leave this separate it gives reviewers opportunity to give their own weighting scheme and it counts twice.
- Or it deletes all the other weights.

- Integrity of the overall submission vs. an outcome measure. Go from numerical scores to a qualitative judgment of the overall application.
- This could be a “bonus,” a place to give credit to those who go the extra mile.
- Wording of the core criterion is preferable to the revision.
- It is a global rating.
- Could be that the results were negative; whatever we found (good or bad) are probably true.
- Magnitude of outcome findings should be weighted more heavily.
- Integrity of research design, outcomes attained, and practicality. Three dimensions which could give overall weighting.
- Breaking out design and outcomes re integrity is a good idea.

## **18. Utility**

- Revision gives new criteria.
- Using research manuscripts to look at utility. Not gathering the kind of information we need to make decisions about practicality.
- Significance is important and does not show up anywhere else. That is a utility component. Significance and contribution vs. practicality and utility. But does that go beyond the intention of NREP? We are not trying to sanction the science.
- There is another option on how to use this material: applicability, generalizability. Could take out the Utility criterion from the raters and have it done by internal staff—whether there has been consumer involvement, whether it is feasible.
- Would group the replication material with effect size and other statistical information.

## **19. Replications**

- Keep for the raters.
- However, would be difficult, and you would want to rate by the design criteria.
- Rather than promoting or creating NREP with an apparent approved list, this tells with a kind of grid that this has reached a certain level of development. With interventions that have had numerous replications, you get at truth in labeling. Could put judgment more in the hands of the prospective users (State or community, not individuals). NREP shows what has promise and is worth looking at; gives tools for the user to see if it would be applicable to their world.

## ***Parking Lot***

- Feasibility
- Real-world capability
- Risks/side effects
- Organizational integration of programs
- Weighting of items
- Representative NREP results
- Ratings
- Significance—overall value, potential contribution to the field

## Discussion

- People say you should not rank programs, but *outcomes*. Want people to send in all replications of their study, with the effects of all the studies in a systematic way. They also said they want to see if there are replications by people who are not the developers.
- Have chosen not to specify outcomes, so you are left with reviewing and summarizing many things.
- Looks like a scientific review panel's report nearly all the way through the review, but then revert to more qualitative measures. SA people may not be able to do it or may have difficulty.
- Have also chosen to label things *promising* or *effective*. Consumer Reports says the label is always problematic because they have to be based on a weighted set.
- Have EBP that are scientifically driven, which might be difficult to combine with the usual SA group.
- The Center for Substance Abuse Treatment (CSAT) and the Substance Abuse and Mental Health Services Administration (SAMHSA) have already picked some criteria that they will fund (abstinence), and we need practices that do that, that are consistent with CSAT *performance indicators*. Politically SAMHSA cannot say "reduced use"; it must say "abstinence." If you specify the outcomes first, you will pay.
- *Feasibility* and *real-world applicability* are concepts we should address.
- *Risk effects* (like side effects) might be important in SA.
- What vulnerabilities are will still be unclear even after we clarify EBP. May end up *overselling* by putting a science label on it. What happens after you take it out of the realm of science? We give an imprimatur to a set of interventions without considering how it might play in a particular area at a particular time with particular people. This process is giving an aura of finality to something in social sciences. Concerned about the expectation for the change in people that NREP may imply. The thing that makes NREP credible is that it has particular criteria, but should not oversell.
- *Skill acquisition* is an important part of EBP, which the program selects.
- *Implementing* NREP programs also needs attention, e.g., being forced to choose an inappropriate intervention for bureaucratic reasons, which means you lose money by making it appropriate
- *Cost feasibility* is among the many categories we cannot evaluate.
- Repeated request for an *operational definition* of EBP; NIDA has declined. But this addresses it.
- *Vulnerability* of consumer comments.
- Some assessments are relevant to the way organizations operate. We should address this.
- Talked about drawing a line beyond which research should not go: should come back to *utility*—research review or other?
- Consumers have multiple *perspectives*, which we must allow room for in their decision for how they want to move on.
- Consumers are likely to interpret *a list* to imply that what is on the list is better than what is out there, that it is approved. Evaluators could provide an effectiveness score, but in addition provide a score evaluating the state of the science—has there been a large-scale community trial that comments on whether it can be implemented. This score would have several components—has it occurred yet? Does it support the intervention? Has it been shown to be

superior? Is there information on the cost of implementation with fidelity? Is there information about the acceptability to consumers and providers?

- Some information can be presented on the crosswalk. There is no place now to score appropriateness. Could condense the above, but not lump everything together.
- Has there been an effort to *evaluate dissemination* into a community? It could be useful. If program dollars are to be made available based on effectiveness, evidence of a community trial would be useful. If it has not been tested, consumers should know that.
- It would be ideal to include Addiction Technology Transfer Centers (ATTCs) as technology transfer.
- Including *training* and implementation is a very high bar now, but should be included in the future.
- The *scoring* the above implies would involve a level of additional complexity.
- In clinical trials you have a Ph.D.-level supervisor; this may not be practical to accomplish with the current workforce.
- Could rate a program as likely effective, but the studies necessary to show effectiveness in the community have not been done.
- Some info routinely collected addresses cost and key aspects.
- *Process*: Once NREP determines a program to be effective, the developer is invited to come to SAMHSA to discuss implementation and dissemination. They receive a template (~9 pages), which includes target population, setting, staffing requirements; this information can be reviewed. Program background info they asks the developer to provide includes: number of recognitions, Institute of Medicine (IOM) classification, length of program, cost, content focus, risk and protective factors, key components and approaches, delivery specifications, essentials of the implementation, proven results, other outcomes, evaluation design. It is consumer driven—complaints to the contrary disprove assertions.
- [Steve Gardner] Start with *efficacy* work of the institutes and move to the... and then to the field. Evaluations get less sophisticated (and less expensive) the farther away from CSAT and SAMHSA you get. Dissemination is increasingly done by commercial developers, who know little about fidelity and adaptation, two issues that remain thorny.
- Sanctioning the *science implies* 2 phases: (1) What criteria determine that an intervention is effective? (2) Implementation, consequences for faulty implementation, etc.
- One challenge of NREP is to keep up with the science. Another is what FDA calls post-marketing surveillance. This is a next-generation challenge that gets into block grant, and efficacy issues.
- A lot of the burden of deciding whether a program can be implemented. Prospective users should be informed that the program has not been attempted, or the degree to which it has been.
- *Adaptability* and ease of *dissemination* is too nuanced for raters to consider. Perhaps a surveillance piece done by different people would take this on.
- Description of *competencies* involved, sophistication of staff, comments from people who have tried the programs, etc would be useful. Now some of this occurs under “cost.”
- We need a place to note core elements or key strategies needed to replicate the program.
- Do series of interventions get evaluated separately or as a package?
- [Dwayne Simpson] *Users’* evaluations—some people use, some do not. Want to know why or why not. Some people download—they want to know who and why, so now they are asking questions of people who download. Trying to get survey responses about whether they

used the material and whether it was successful implemented. Sometimes people have implemented their material and are proponents of it, but they (the developer of the material) knew nothing about it.

- [Paul Brounstein] *NREP was developed* as the first piece of an entire system. Identifies EB, next is development, next is implementation, next dissemination. Northrop Grumman has been trying to collect lessons learned. It is important to identify how these materials are really used. Since there is no money for site visits, they have to rely on what the developer tells them. A lot of information is available in text but not electronically. Would love to know what people's experience with these programs is. Also do not know to what extent people think the "solution" is just a large pain. SAMHSA has developed four funding mechanisms that will do the things many at the table are concerned about, but the question remains whether the organizational will exists to elicit this information and results. The mechanism is in place, but whether it will be continued next year is moot.
- Raters should be commenting by outcome, by population, by significance. No intervention is effective for everyone everywhere. Finney and Prendergast reviewed 1,100 interventions—none had power to look at gender by three different ethnic groups.
- Discrimination between scientific content and ratings is one thing. Model status will be a whole different review. Will need a review group comprised of consumers (i.e., buyers, not patients). Training is significant and is a cost that will have to be born by people who want to have model status. People who get downloads can pay or do reports to elicit post-marketing status.
- We are talking about who is walking up the stairs and why. In MH treatment, States were very willing to test SAMHSA's tool kits. SAMHSA hope to use some grant mechanisms for this purpose. There is a lot of interest in things that could happen.
- There is some level of responsibility for the one who wants to further the product.
- Consumer Reports thinks it is extremely important to say that there is *no information available*.
- [Gardner] The watchword in creating those matrices is that the data have to specific to that population (e.g., slide 19).
- People are taking EBP and implementing part of them. There needs to be an understanding of what the crucial pieces (*core components*) are. NREP could identify core elements with ratings. NREP has arranged a series of conference calls with developer and implementer around a program, which is transcribed and on the Web—includes difficulties and solutions.
- *What is a program?* Is a TCU module a program? Some groups just want to be on the list even though they never disseminate anything. A combination of therapies is not a program, but a practice. May be the burden of the submitter; the intervention would be evaluated according the submitter's description. Organization and integrity of program. If submit a single intervention, what is the context in which it is being delivered? Must include the context in which an intervention is delivered (in a court, in a home, in a hospital, etc.). These issues were not relevant when CSAT developed NREP; it was what CSAT wanted to promote.
- Must show fidelity measures and that you are getting the same outcomes. But a program would not have to recreate a clinical trial. There has to be a mechanism to tell the scientists how it fits.
- *Weighting and significance:* The two are related. As a reviewer it is important what value the initiative brings to the field. Some categories are critical; others are nice but not necessary for

approval. Do we want room for a global score on significance and the importance of the contribution? NREP has teams of rotating people (taking into account substantive limitations) who review. Weighting issue is dealt with by providing NREP criteria in an effort to get feedback and to guide. But the way raters attach significance to particular items is up to the individual scientist. They have three reviewers and average the scores to come up with one conclusion. We are weighting whether intentionally or not. The format is already set up for weightings, but the individual item responses were not included. The format lends itself to a quantitative approach and the greater precision the better. We could group criteria. There already is a lot of interpretation built into each item.

- Even if you work with clusters, do you still want to arrive at a global score? Do not need it if you stick to a *two-tiered evaluation*: part 1: scientific, quantitative, objective; part 2: interpretative. It is possible to make reasonably accurate scientific estimations of the criteria; then remove the scientists.
- The problem with scientific weighting, is that we would need to structure the weights in such a way that they reflect our overall judgments—what do we consider to be good science? Then a make subjective judgment about how important these things are to our overall judgment. Would be useful to provide broader categories for people to score other than what is in the current document.
- This requires a small *work group* to prepare something on a weighting of scientific merit areas. One issue is the weighting of the items (the entire group would be involved); and figuring out what the clusters might be (small group), which could be included as a scale score = integrity.
- We should wait because at this point we do not know what the criteria will be. Once the criteria have been modified, this would be a next step.
- *Adoption and implementation* are not automatic. Once assessments are on-line, implementers can compare their results with others' results. A larger concern is the cross-currents: programs vs. stages of treatment. We have to leave room for people who are assembling their treatment programs as they go along. This is a shift from what has been done in prevention. The word program has been taken in the SA field. There must be components designed for patients at various stages. There's an important role for a guidance document (*user's manual*) for using a system like this—Where does what I do fit into this box?
- *Usability* must be dealt with more inclusively.
- Drug court may come in as an intervention and not a program.
- Categorization of interventions by the *stage* at which they occur should be considered to give a definable package of goods that would be exportable. Interventions could be laid in at various phases. This would allow people to submit some testable component that is not the whole program.
- If it is written right, it could be educational.
- However, treatment is paid for by specific intervention, e.g., days of methadone care.
- The developer needs to identify the level of interest.
- [Brounstein] For NREP *outcomes* had to be at least theoretically relevant to SA. Often you have to look at proxies that relate to SA (e.g., harsh punishment); as a reviewer you would have to use subjective weighting. Rely on reviewers to assign importance to the outcomes studied. This is a system. On the back end of this are the science academies. When we can get SA outcomes we do, but when it is not appropriate for the population, we adapt.
- Some of this could be incorporated into a guidance document.

- What is going to be classed as effective is a long string of dissimilar interventions. It will be a wide range. Could group by outcome at each stage.
- It was a decision by the group to not pre-specify outcomes. However, Dwayne's is the closest we have to work in the field, and that involves *staged outcomes*. So, inevitably it may have to come to this.

### ***Wrap-Up***

- Where will this go? HSRI will summarize comments. A transcript of the meeting will become available. The first criterion theory gave an easy answer, but for many there was not. SAMHSA will have to determine what needs more small-group discussion. SAMHSA will respond to the comments of the group (e.g., utility and dissemination). Would be helpful to have a summary of the overriding concerns that need to be addressed, e.g., guidance document and what should be included in it. Set up a listserv for the MH groups, and HSRI will do it for this group also.
- Simpson offered compliments to everyone who has worked on this—we are not talking about a major makeover, just nuances.
- NREP has a lot of work to do subsequent to this meeting. The first portfolio will be the adolescent portfolio. CSAT put 2003 funds into the existing NREP contract; they are expected to begin work this year. At the same time a new contract is in the works, so there is a big discussion about transition. The short-term (2 or 3 months post-meeting) issue is guidance around terminology, endorsement of criteria, etc.
- IOM information was incorporated. How do we look at these criteria across all three centers? This group has focused on scientific merit, while others seem to have focused on the consumer. A group is in place to talk to people in all three groups to discuss common and diverse issues.
- The NREP process—review of science and dissemination—would be a wonderful model process for science to service and develops dissemination application. The Federal responses to those two reports are encapsulated in NREP.

### ***National Cancer Institute (NCI)'s Use of NREP***

Dr. Schinke explained the Northrop Grumman Web site Model Programs home page.

NCI's NREP, available as one of the model programs, at Cancer Control PLANET. Topics available include: 5 a day (fruits and vegetables), breast cancer, cervical cancer, physical activity, and sun safety.

The group interventions by category and then list them on the Web. Programs can be searched by various factors (age, race/ethnicity, setting). Hyperlinks connect to other aspects of the intervention. Material reviewed is rated by integrity, utility, replications, and cultural sensitivity. They include only reviewed articles, whereas SAMHSA reviews whatever material it receives. Distribution for this and the SAMHSA and CSAT programs are academics and community organizations. Nearly all disseminate their materials through various for-profit organizations as well as foundations. Includes information to download or order materials having to do with the

programs, including program adaptation guidelines. All materials are for providers. (The money for this came from Cancer Prevention and Control.)